Preprint
UCRL-JC-148363-REV-1

Middle East, North Africa, and Western Eurasia Seismic Research Database

J.L. O'Boyle, S.D. Ruppert, T.F. Hauk, D.A. Dodge, M.D. Ganzberger, F. Ryall

This article was submitted to Seismic Research Review, Tucson, AZ, September 23-25, 2003

U.S. Department of Energy

Lawrence Livermore National Laboratory

July 14, 2003

Approved for public release; further dissemination unlimited

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

This work was performed under the auspices of the United States Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

This report has been reproduced directly from the best available copy.

Available electronically at http://www.doc.gov/bridge

Available for a processing fee to U.S. Department of Energy
And its contractors in paper from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-mail: reports@adonis.osti.gov

Available for the sale to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/ordering.htm

OR

Lawrence Livermore National Laboratory
Technical Information Department's Digital Library
http://www.llnl.gov/tid/Library.html

UCRL-JC-148363-REV-1

MIDDLE EAST, NORTH AFRICA AND WESTERN EURASIA SEISMIC RESEARCH DATABASE

Jennifer L. O'Boyle, Stanley D. Ruppert, Teresa F. Hauk, Douglas A. Dodge,
Michael D. Ganzberger, and Flori Ryall

Lawrence Livermore National Laboratory

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation

Contract No. W-7405-ENG-48

ABSTRACT

The National Nuclear Security Administration (NNSA) Ground-Based Nuclear Explosion Monitoring Research and Engineering (GNEM R&E) Program at Lawrence Livermore National Laboratory (LLNL) has made significant progress populating a comprehensive Seismic Research Database (SRDB) used for deriving seismic calibration parameters for the Middle East, North Africa and Western Eurasia (ME/NA/WE) regions. In addition to an overview of select individual information products, we present an overview of our visualization, integration, validation, and organizational processes. Development of these processes and the LLNL SRDB was necessitated by both the very large amount of data and information involved (over 15 terabytes) and the varied data and research result formats utilized. The LLNL SRDB allows for the collection of raw and contextual seismic data used in research, provides an interface for researchers to access data, provides a framework to store research results and integrate datasets, and supports assembly, integration and dissemination of datasets to the NNSA Knowledge Base (KB).

The LLNL SRDB is a flexible and extensible framework consisting of a relational database (RDB), Geographical Information System (GIS), and associated product/data visualization and data management tools. This framework is designed to accommodate large volumes of data in diverse formats from many sources (both in-house-derived research and integrated contractor products), in addition to maintaining detailed quality control and metadata. In order to efficiently organize information within the LLNL SRDB, it was necessary to automate procedures needed to create and update database tables, but a large effort is still required by technicians and scientists to load special datasets, review results of automated processing and resolve quality issues. The LLNL SRDB currently has 3 million reconciled event origins and arrivals from several global, regional and local seismic bulletins and 30 million waveforms from 570,000 seismic events. We also use the LLNL SRDB for integration and have developed procedures, algorithms, and tools, organized as a tool set called KBITS, which permits datasets to be merged and dataset quality control to be maintained. KBITS can be used to integrate waveform and bulletin data into unified datasets. This suite of tools is now being used by LLNL for content development and by Sandia National Laboratory to assist in creation of the KB. It enables inclusion of data from contractors, universities, and NNSA collaborators for content development and to provide coherent, reconciled and integrated results to the NNSA KB. The SRDB integration framework also supports and facilitates NNSA ROA and BAA projects.

The LLNL SRDB provides the basis for synergistic development of all LLNL GNEM R&E Program research. It is not only a coherent framework in which to store and organize very large volumes of collected seismic waveforms, associated event parameter information, and spatial contextual data, but also provides an efficient data processing/research environment for deriving a wide range of information products required to support the ME/NA/WE regionalization program. Products assembled, integrated and validated using the LLNL SRDB are grouped into five main categories: reference and contextual information, detection data, calibration and ground-truth data, event location products and event identification products. Using the LLNL SRDB, we are combining travel-time observations, event characterization studies, and regional tectonic models to assemble a library of ground-truth information and phenomenology (e.g. travel-time and amplitude) correction surfaces. Corrections and parameters distilled from the LLNL SRDB provide essential contributions to the NNSA KB for the ME/NA/WE region and facilitate calibration of seismic monitoring stations, thereby improving capabilities for underground nuclear explosion monitoring.

OBJECTIVE

The primary objective of the LLNL Seismic Research Database (SRDB) is to facilitate development of information products for the Ground-Based Nuclear Explosion Monitoring Research and Engineering (GNEM R&E) Middle East, North Africa and Western Eurasia (ME/NA/WE) regionalization program. The LLNL SRDB provides efficient access to, and organization of, millions of seismic events and associated waveforms, while also providing the framework to store, organize, integrate and disseminate information products for delivery into the National Nuclear Security Administration Knowledge Base (NNSA KB).

The LLNL SRDB provides a unified framework for all seismic data and information products as outlined in Figure 1. This requires the reconciliation and merging of data derived from different sources and methods and of varying quality, along with the ranking of the data sets based on relative quality. Integration occurs on many levels in order to generate data, data sets, databases or knowledge bases. The LLNL SRDB also allows for the integration of research with contractor research products and other available data sets. In order to efficiently organize information within the LLNL SRDB, it was necessary to automate procedures needed to create and update database tables, but a large effort is still required by technicians and scientists to load special data sets, review results of automated processing and resolve quality issues. Sufficient metadata (including measurement procedures, codes, comments and measurement errors) are stored at each step in the data creation and analysis process to allow re-creation or verification of results at any stage in the processing flow.

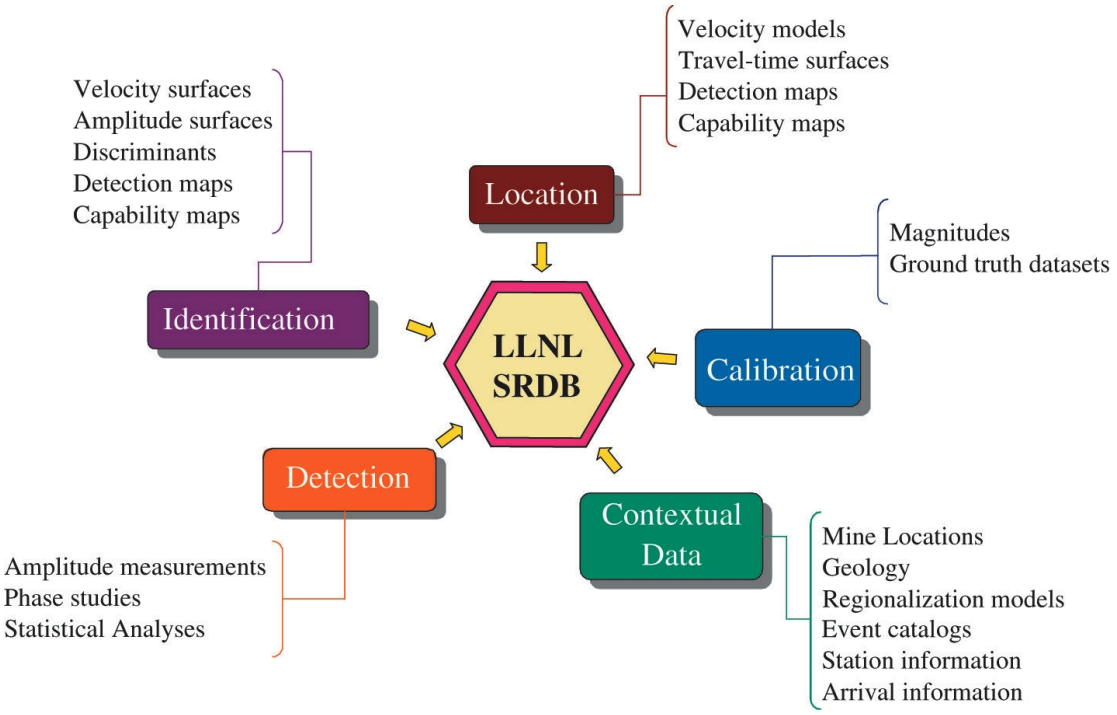

Figure 1. The LLNL SRDB provides a unified framework for contextual/reference data and information products. The LLNL SRDB provides efficient access to, and organization of, thousands of seismic events and associated waveforms, while also providing the framework to store, organize, integrate and disseminate information products for delivery into the National Nuclear Security Administration Knowledge Base (NNSA KB).

Information products created using the LLNL SRDB may be grouped under two major categories: primary data products and derived products. The primary products are those developed in the process of collecting the raw materials for calibration: ground truth data, waveform data, event catalogs, phase pick information, regional station information and instrument responses. The derived products (distilled from the organized raw

seismological data) are models and corrections that improve detection, location and identification functions. In order to calibrate seismic monitoring stations, the LLNL SRDB must incorporate and organize the following categories of primary and derived measurements, data and metadata:

Contextual and Raw Data
1. Station Parameters and Instrument Responses
2. Global and Regional Earthquake Catalogs
3. Selected Calibration Events
4. Event Waveform Data
5. Geologic/Geophysical Data sets
6. Geophysical Background Model

Measurements and Research Results
1. Phase Picks
2. Travel-time and Velocity Models
3. Rayleigh and Love Surface Wave Group Velocity Measurements
4. Phase Amplitude Measurements and Magnitude Calibrations
5. Detection and Discrimination Parameters

Corrections and parameters distilled from the LLNL SRDB provide needed contributions to the NNSA KB for the ME/NA/WE region and will improve capabilities for underground nuclear explosion monitoring. The contributions support critical functions in detection, location, feature extraction, discrimination, and analyst review. Figure 2 outlines the processes of data collection, research and integration that result in contributions to the NNSA KB.

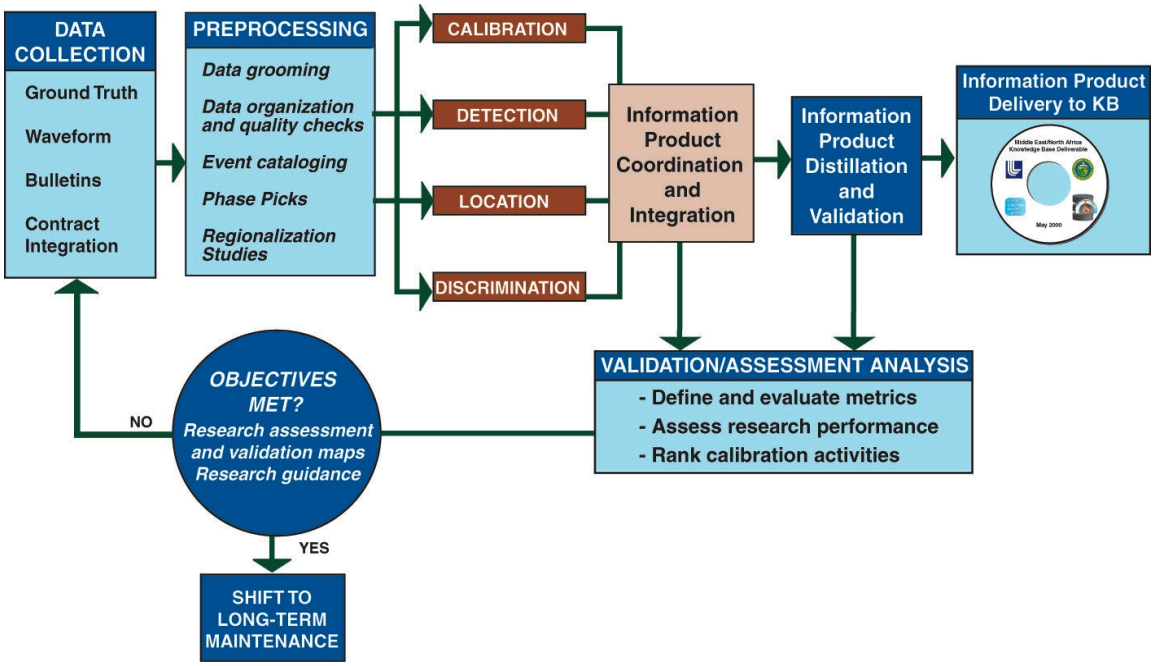

Figure 2. Seismic Research Database (LLNL SRDB) process flow. The LLNL SRDB is the framework that supports development of information products related to seismic calibration research. Integration and synergy of data and research and visualization tools are made possible through the LLNL SRDB.

RESEARCH ACCOMPLISHED

The LLNL SRDB is a framework consisting of an ORACLE relational database (RDB), Geographical Information System (GIS), and associated product/data visualization and data management tools. The LLNL SRDB is necessary for the storage and organization of very large volumes of collected seismic waveforms, associated event parameter information, and spatial contextual data. It also provides an efficient data processing/research environment for deriving location and discrimination correction surfaces. Figure 3 outlines the interaction between all components of the LLNL SRDB. The relational database is organized in CSS3.0 (Center for Seismic Studies Version 3.0) format with custom extensions. This format provides parameter defined tables for different elements of seismic data, such as event and station information, as well as allowing for customized tables to be developed for specific research needs or results and the compatibility with other organizations.

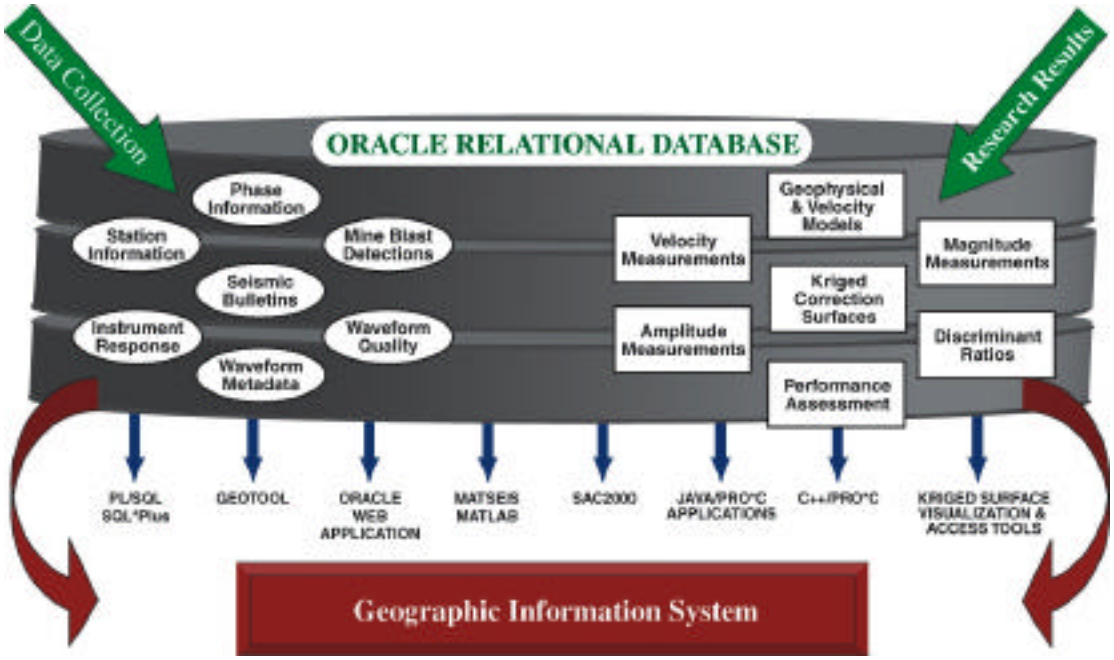

Figure 3. The LLNL SRDB integrates a relational database, a Geographic Information System (GIS) and visualization/data management processes. Many aspects of GNEM R&E are contained within the database and can be accessed using a wide variety of tools.

The first tier of the LLNL SRDB development is largely the collection of catalogs and waveform data, parsing, conversion and extraction of data and database entry. What makes these tasks challenging are the multitude of catalog and waveform data formats, and the sheer volume of the data. LLNL has succeeded in automating many of the collection, parsing, reconciliation and extraction tasks, individually. We are currently working towards a more integrated system so that much of this first tier of the LLNL SRDB development will be almost completely facilitated by automated processes.

Currently, the automated procedures are organized into a tool set called KBITS, which uses modern object-oriented and database programming practices to combine a number of tools and libraries that incorporate automation where possible to streamline the data set loading, reconciliation, and integration processes. This framework also provides the primary mechanism for merging and reconciling derived calibration data from contractors, individual researchers, and NNSA laboratories into a single product. There are six main classes of products that require integration effort and four main reconciliation issues for each class summarized below:

Data set Class	Reconciliation Issue
Core tables / Core files (e.g. velocity models) Catalogs	Initial merge Updating

Waveforms Derived products (internal and external) Spatial/Contextual information Assembled databases / Knowledge bases	Validation Back propagation / synchronization

To assist us in the integration and reconciliation tasks related to calibration research, we have written a set of 130 C++ classes designed primarily as components of computer programs for use by the GNEM R&E program. A class is a software abstraction of an entity from the problem domain, which combines all of the attributes and actions of a single entity into one data type. This allows programmers to work at a higher level of abstraction, thus improving productivity and software maintainability. Of the classes we designed, about 10 are service classes with no specific seismological capability, *e.g.* time class, architecture class, etc. All the remaining classes are specialized to some degree for use in seismological applications. All of the collection classes have database-aware specializations as well, since most of the programs developed with the classes interact extensively with the LLNL SRDB database. The class library has built-in support for the CSS3.0, SAC, GSE2.0, and PC-SUDS data formats. Using the classes has proven to be an effective means of building the utility programs required to support the LLNL SRDB as described in the following paragraphs.

We have developed an automated method/tool (Orloader) to load various seismic bulletins of different formats into native-format database tables, which retain many of the fields provided in the original bulletin, as well as assign a unique origin identifier (*orid*) to each event. Individual bulletins are then merged into a single CSS3.0 **origin** table retaining the author of the original bulletin. The native tables are linked to the origin and event tables created by Orloader through an remap table. Using the remap table, any origin in the target CSS3.0 table can be uniquely associated with the corresponding native origin. Because the same event may often be reported by different organizations, there will be instances in which new origins correspond to origins that are already in the database. A space-time correlation algorithm is used to reconcile events between individual catalogs. Events common to multiple bulletins will be assigned common event identifiers (*evid*) while retaining unique *orid*s. The **event** table stores the preferred origin for each event based on a ranked list of catalog preferences. Phase arrivals provided with native catalogs are loaded into the **assoc** and **arrival** tables.

To generate the event waveforms for the LLNL SRDB, we have developed automated methods/tools (Ddload, UpdateMrg) for extracting event waveforms from continuous data and inserting them into the appropriate database tables. Event selection can be restricted by time, location, event magnitude or a specific list of *evids*. Ddload extracts waveforms from continuous data in the native format, converts the data to CSS3.0 format and creates entries in staging tables (**wfdisc**, **wftag**). The **wfdisc** table contains pointers to the waveform files and the **wftag** table associates the waveform file with the *evid* used in the extraction. Currently, we can perform waveform extraction on data in SAC, SEED, CSS, SEGY and GSE formats. The UpdateMrg program is used to insert the staged **wfdisc** and **wftag** records into the target tables. During the insertion process, UpdateMrg checks waveforms for poor or missing signals, which are then recorded in a metadata table. If waveforms for a specific event are segmented, the segments will be merged together; if a gap exists between segments, the space will be filled with zero-value data points. If two segments that are being inserted overlap or a segment for an overlapping time period already exists in the target tables, a correlation test is done before the two segments are joined.

Also included in KBITS are a suite of consistency tests and related updating/repair procedures (DBcheck / DBfix) to assist both the data set loading staff and researchers in maintaining the highest quality collection of data possible. These test procedures consist of PL/SQL stored procedures used for identifying and resolving inconsistencies between data sets (core tables and measurement tables) to be integrated as well as maintaining consistency in the final LLNL SRDB and products to be delivered.

The second tier of LLNL SRDB development activities, the development of correction surfaces and other calibrations, is less susceptible to complete automation as these activities require the judgement of scientists skilled in the interpretation of often highly unpredictable event observations. However, development of proper tools to extract observations and parameter data from the LLNL SRDB and to make thousands to millions of waveform measurements can significantly increase the efficiency of the scientists who construct and validated integrated calibration surfaces. This trend is likely to continue and even accelerate leading ultimately to larger investments in enabling technologies offset by declining expenditures on the routine construction of calibrations. The first tool to be developed is a prototype body wave amplitude measurement tool. The tool is written in Java and incorporates a GUI and GIS system for project management and event selection with a semi-

automated, interactive measurement window picker. Development of this tool has produced a number of reusable classes that we expect will be helpful in further automation efforts.

The LLNL SRDB also provides the framework to perform end-to-end process validation and testing of research results. Validation and assessment tasks include the definition and evaluation of specific metrics, analysis and use of metrics to define the stopping criteria for particular location and identification tasks, and ranking and performance evaluation of calibration activities. Success and efficient implementation of validation and assessment tasks require a framework that ties research results to data and measurements used to create each calibration product and that also supports statistical and visualization tools for performing the validation tasks. The metadata inherent in the LLNL SRDB allow multiple realizations of calibration products, created with different processes and assumptions, to be created, stored, and compared with one another. Thus the LLNL SRDB facilitates the distillation, documentation, and delivery of information products.

Contextual and Raw Data

Station Parameters

Seismic station information is a metadata requirement needed to support all stages of seismic waveform analysis. These metadata include such parameters as station operation dates, location and elevation, type of channels and instruments, sampling rates, and instrument responses. Our main source of this information is Incorporated Research Institutes for Seismology (IRIS) "dataless" SEED files, which are provided by each of the networks affiliated with IRIS. Other station information has been obtained through Internet station books and AutoDRM systems. CSS3.0 **site** and **sitechan** table entries (listing station location, available channels, sensor orientations, operation dates, etc.) were created for almost all IRIS affiliated networks as well as many other stations with waveforms in the LLNL SRDB. All station parameters are reviewed among existing information sources, and conflicts must be resolved between sources and in reference to waveforms available before database tables are updated. Over 2400 station and array element table entries have been updated, but reliable parameter data do not exist for some stations if only minimal or inconsistent information is available. **Instrument** and **sensor** tables are used to document instrument type and response for each station and channel. The Seismic Analysis Code (SAC) has been modified to interpret response information in RESP or FAP format for use in performing instrument response corrections on waveform data by using the EVALRESP software library available from IRIS.

Event Bulletins

Reference event locations and origin time information are necessary in most stages of our seismic processing and research. Bulletin information from many global, local, and regional earthquake catalogs has been incorporated into the LLNL SRDB and provides a much larger source for event selection. This combined and reconciled source facilitates comparison of event parameters provided by multiple networks with different degrees of location accuracy and provides a wider range of magnitudes and event types. The global catalogs include: United States Geological Survey (USGS) Monthly (Final) Preliminary Determination of Epicenters (PDE) catalog and Earthquake Data Report (EDR) catalog, Bulletin of the International Seismological Centre (ISC), Harvard Centroid Moment Tensor (CMT) catalog, and the International Data Centre Reviewed Event Bulletin (REB). We have also compiled numerous regional and local catalogs for the ME/NA/WE region, including Jordan, Israel, Scandinavia/European Arctic and the Kola Peninsula. Figure 5 shows a map of the event location coverage provided by several of the bulletins in the ME/NA/WE region. We have established a number of collaborative agreements with countries and institutes in our study region to obtain both local seismic catalogs and ground truth information as well as seismic waveform data.

Waveforms

We are collecting seismic data from IMS primary and auxiliary stations, as well as surrogate stations (for IMS stations not yet installed) and other stations needed to support calibration in the region of study. We have obtained over 10 years of continuous data for important ME/NA/WE stations from IRIS, Institut de Physique du Globe de Paris GEOSCOPE program, the GEO-Forschungs Zentrum/Potsdam, Germany GEOFON program and other data centers. Data for particular events has been obtained from the prototype International Data Center (PIDC) and the United States National Data Center (US NDC). The Center for Monitoring Research (CMR) and NORSAR have provided waveforms for special regions/events, such as the Novaya Zemlya test site. Stations in the ME/NA/WE region with event waveforms stored in the LLNL SRDB are shown in Figure 4.

Most of the archived waveforms in the LLNL SRDB are from events located within the ME/NA/WE region and occurring between 1976 and 2002. Although the continuous data remain archived on tapes, data segments for seismic events are extracted from the continuous waveforms. The number of waveforms in the LLNL SRDB is now almost 30 million, which represents 570,000 seismic events. The events with waveforms in the LLNL SRDB are shown in Figure 5.

Other Reference and Ground Truth Data sets

Projects to develop reference data sets include waveform correlation and subspace detectors to provide statistics on mining activity and a regionalization model based on *a priori* geophysical knowledge used as a reference model for model-based correction surfaces. Collaborations with Cornell University, USGS and other organizations allow us to obtain reference data sets useful as background and supporting information for research. Most of these data sets are provided as Geographic Information System (GIS) products, which allow us to integrate them with our data sets. The data sets include geographic, geopolitical, mining industry, geologic and tectonic information for the ME/NA/WE region. We also collect photographs and satellite imagery to support ground truth determination.

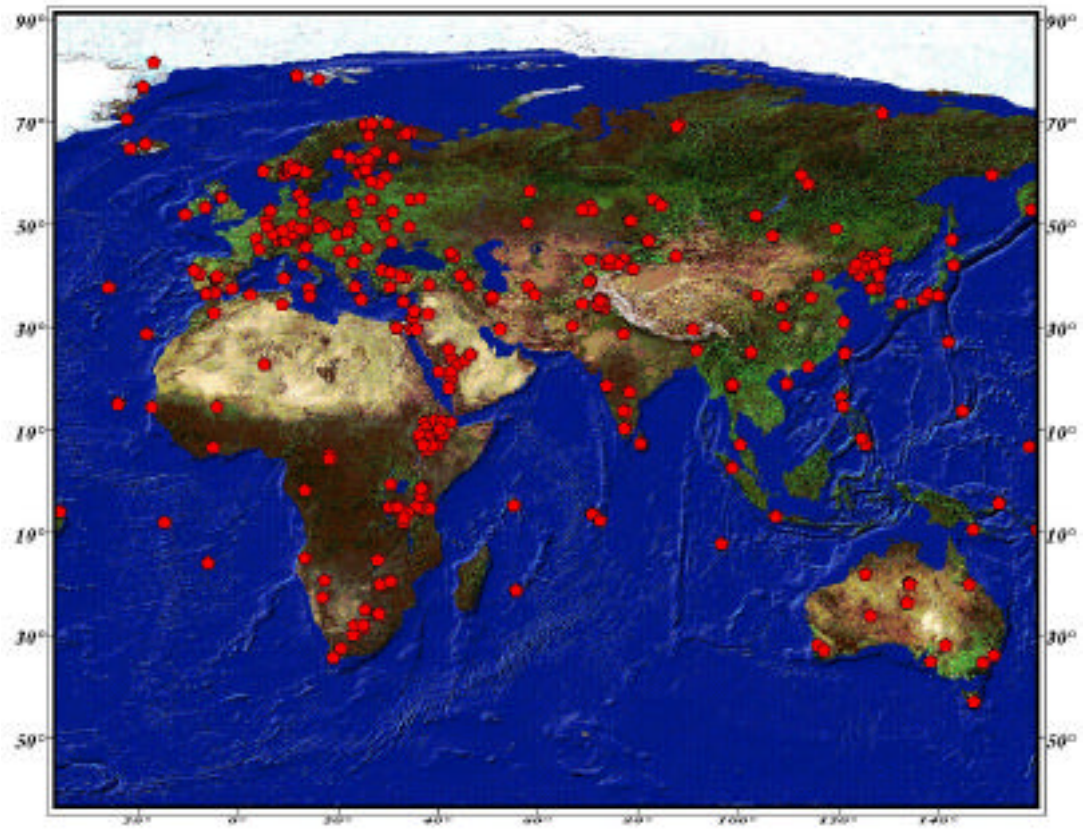

Figure 4. Seismic stations in the ME/NA/WE region with event waveforms stored in the LLNL SRDB.

Measurements and Research

Phase Information

LLNL researchers and analysts have made phase picks for over 1700 events to yield over 24,500 travel-time observations available to the research team for location and discrimination projects. Phase information is recorded in the **arrival** and **assoc** tables, along with pick and waveform quality and other comments that are recorded in a **remark** table. Augmenting the picks, we added 36 million ISC phase arrival measurements, 2 million REB arrivals and 12 million EDR arrivals along with picks from several regional bulletins to be used for travel-time correction studies.

Seismic Location Research

Seismic location researchers utilize event catalogs (especially ground truth), waveforms and phase picks to develop travel-time corrections. Improvement in seismic location is accomplished by combining model-based and empirical travel-time corrections. After a calibration model is applied, empirical corrections are calculated using Modified Bayesian Kriging algorithm with travel-time residuals for suitably well located calibration events. The travel-time and velocity models developed through this research are an important part of the LLNL SRDB. See the LLNL Location and Detection Research (Myers *et al*., 2003 this Proceedings) paper for more details about these aspects of the GNEM R&E research.

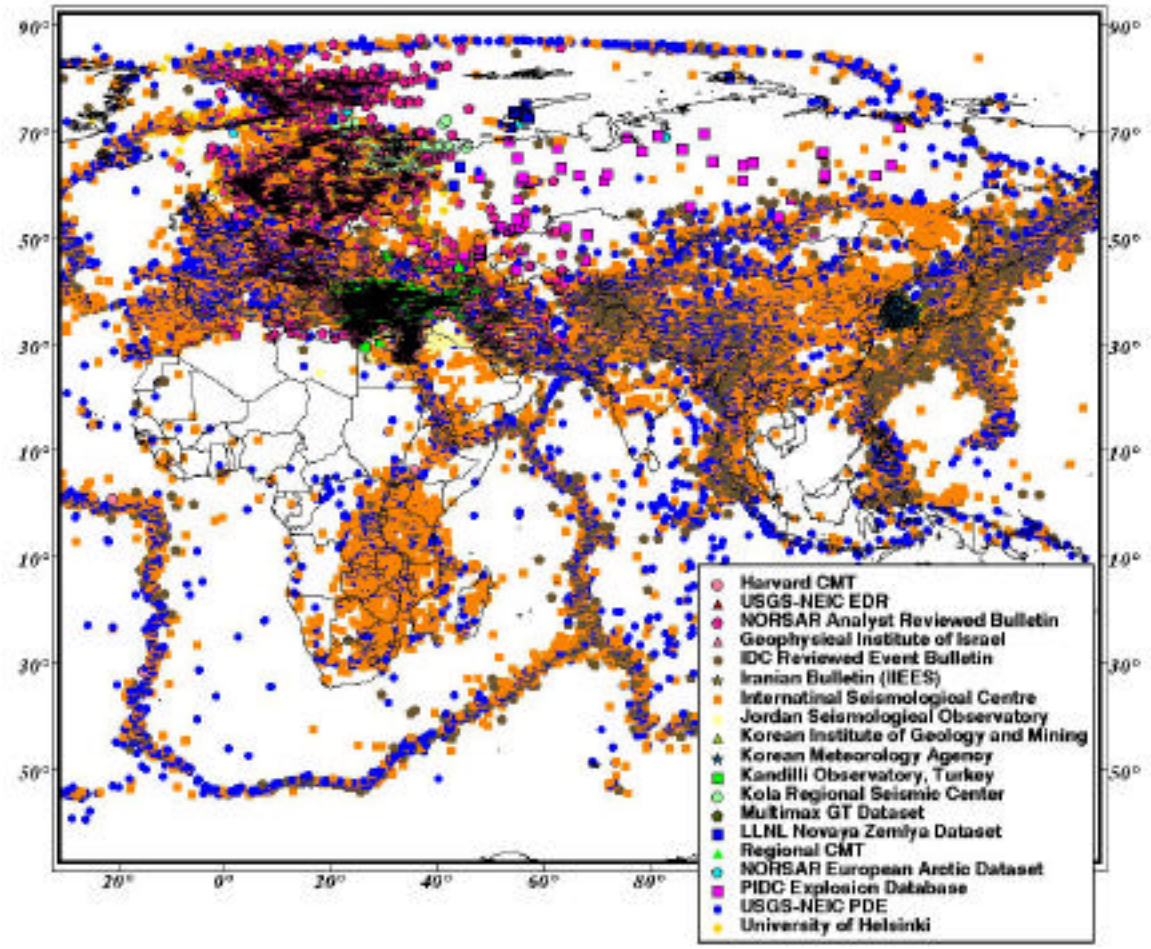

Figure 5. Plot of seismic events in the ME/NA/WE region with waveforms in the SRDB. The LLNL SRDB contains multiple seismic event catalogs that have been reconciled into one database table. These catalogs range in scope from global to regional to special ground truth data sets and provide a much broader range of event magnitudes and event types than any single catalog.

Event Identification and Magnitude Research

The event identification researchers also use event catalogs, waveforms and phase picks to develop products that provide seismological information to improve the ability to identify underground nuclear explosions, by discriminating them from the background of earthquakes. The magnitude research results in calibration products and magnitudes for 1-Hz regional phases, broadband coda waves, and long-period surface waves. This provides the necessary calibration information for regional data in order to calculate seismic magnitudes for use by nuclear monitoring functions (detection, location, identification, and characterization). Custom tables in the

LLNL SRDB database store measurements and related parameters for surface wave group velocities, body wave amplitudes and magnitudes. See the Regional Body-Wave Corrections and Surface-Wave Tomography Models to Improve Discrimination (Walter *et al*., 2003 this Proceedings) paper for more details about these aspects of the GNEM R&E research.

Data Access

Different researcher needs for data and metadata require that subsets of data be provided in a format easily accessible to many diverse types of software and analysis tools. Therefore, the LLNL SRDB access tools have been designed to utilize the power of the relational database to facilitate efficient queries and data retrieval. The Seismic Analysis Code (SAC) software used by researchers provides direct access to database table information and waveforms and uses the response files to perform instrument response corrections. For spatial queries and organization, we have adopted a Geographic Information System (GIS), which provides a framework to store and manipulate spatially defined data. The GIS is linked with the ORACLE database to provide joint spatial and relational queries. We have integrated a large number of our research and contextual data sets into the GIS products in conjunction with other data sets, including contractor products. Researchers can use the GIS to browse many of their research data sets combined with integrated data sets, and also perform basic comparisons, queries, and analyses.

Given the large quantity of data now managed by the LLNL SRDB, the need arose to create the efficient "production" level seismic data selection, processing and visualization tools necessary to meet programmatic and NNSA KB needs. These tools, along with data browsers to allow visualization and quick access to both data and delivered information products, were developed in collaboration with Sandia National Laboratories. The tools provide such functionality as deriving location corrections, making amplitude and magnitude measurements, and developing discriminants.

CONCLUSIONS AND RECOMMENDATIONS

We have made a major effort to provide a product development environment that encourages the natural synergies among each of the separate research efforts of the GNEM R&E researchers. The derivation of reference and ground truth data sets and location and identification products takes place in an integrated environment with changes and improvements in one area being used to facilitate development of the remaining areas. We utilize the LLNL SRDB as an integrating framework to provide the basis for synergistic development of all GNEM R&E. By combining travel-time observations, event characterization studies, and regional wave-propagation studies for ground truth and regional events, we have assembled a substantial library of ground truth information (origin times, locations, depths, magnitudes), mine explosion statistics, tomographic models and travel-time and body-wave correction surfaces.

The LLNL SRDB allows for the collection of raw and contextual seismic data to be used in research, provides an interface for the researchers to access data, provides a framework to store research results and integrate data sets, and supports assembly and dissemination of data sets to the NNSA KB. A wide range of information products required to support the ME/NA/WE regionalization program are being derived from waveforms, station parameters, and event origin information contained in the LLNL SRDB. Corrections and parameters derived, assembled, integrated and validated using the LLNL SRDB provide essential contributions to the NNSA Knowledge Base for the ME/NA/WE region and will improve capabilities for underground nuclear explosion monitoring by supporting critical functions in detection, location, feature extraction, identification, and analyst review in the Middle East, North Africa and Western Eurasia.

REFERENCES

Myers, S., D. Harris, M. Anderson, W. Walter, M. Flanagan, F. Ryall (2003), LLNL Location and Detection Research, 25th Annual DTRA/NNSA Seismic Research Review Proceedings.

Walter, W., M. Pasyanos, A. Rodgers, K. Mayeda, A. Sicherman (2003), Regional Body-Wave Corrections and Surface-Wave Tomography Models to Improve Discrimination, 25th Annual DTRA/NNSA Seismic Research Review Proceedings.

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.